Course Title: Artificial Intelligence **Sheet #2**

1. Consider the problem of predicting how well a student does in her second year of college/university, given how well they did in their first year. Specifically, let x be equal to the number of "A" grades (including A-. A and A+ grades) that a student receives in their first year of college (freshmen year). We would like to predict the value of y, which we define as the number of "A" grades they get in their second year.

   Use the following training set of a small sample of different students' performances. Here each row is one training example.

x	y
5	4
3	4
0	1
4	3

Recall that in linear regression, the hypothesis is

$$h_\theta(x) = \theta_0 + \theta_1 x$$

, and we use *m* to denote the number of training examples.

a) For the training set given above, what is the value of *m*?

b) Recall the definition of the cost function What is *J(0, 1)*?

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^{m} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

c) Suppose we set $\theta_0=0$ and $\theta_1=1.5$. What is $h_\theta(2)$?

2. Suppose you have *m*=14 training examples with *n*=3 features (excluding the additional all-ones feature for the intercept term, which you should add). The normal equation is

$$\theta = (X^T X)^{-1} X^T y$$

For the given values of *m* and *n*, what are the dimensions of θ, X, and y in this equation?

Use Normal Equation to solve for θ

3. Suppose you have a dataset with *m*=1000000 examples and *n*=15 features for each example. You want to use multivariate linear regression to fit the parameters to data. Should you prefer gradient descent or the normal equation? Why?

4. Suppose you have a dataset with *m*=50 examples and *n*=200000 features for each example. You want to use multivariate linear regression to fit the parameters to data. Should you prefer gradient descent or the normal equation? Why?

5. Suppose students *m*=4 have taken some class, and the class had a midterm exam and a final exam. You have collected a dataset of their scores on the two exams, which is as follows:

midterm exam	(midterm exam)^2	final exam
89	7921	96
72	5184	74
94	8836	87
69	4761	78

You'd like to use polynomial regression to predict a student's final exam score from their midterm exam score. Concretely, suppose you want to fit a model of the form $h_\theta(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$, where x_1 is the midterm score and x_2 is (midterm score)^2. Further, you plan to use both feature scaling (dividing by the "max-min", or range, of a feature) and mean normalization.

What is the normalized feature $x_2^{(2)}$? (Hint: midterm = 89, final = 96 is training example 1.)

Best wishes

Dr. Sherin El Gokhy